To quickly recap, we began this project to further investigate our hypotheses about the association between various factors (such as holidays, weather, and contributing factors) and the number of car accidents.

For holidays, we wanted to find the holiday with the most number of accidents, which turned out to be Veteran's day (tag Query 1). We also hypothesized that since holidays often meant parties involving alcohol, there would be a higher rate of accidents in months with more holidays. However, from the query tagged Query 2, it appears that there is no significant evidence in the data to support this claim.

Furthermore, we aimed to investigate various weather influences on accident rates. In our study, we chose to specifically study three factors: precipitation levels, average temperature, and wind speed.
In general, we hypothesized that the association with number and severity of crashes was as follows: significant and directly proportional for precipitation levels, significant and inversely proportional for temperature, and slightly directly proportional for wind speed. For temperature (Query 3), we found that extremely low temperatures was associated with an increase in severity of crashes. The frequency of crashes tended to decrease as temperatures went up, but increased as the temperature increased to the > 75F range. We presumed that this was due to more activity occurring in the summer months, but more testing and tables would be required to further investigate this. Next, we found that lower precipitation levels had the highest number of average crashes per precipitation level, and no major accidents occurred with a precipitation level of 5 inches (Query 4). Lastly, we found that there was a significantly higher proportion of accidents that involved an injury or death when wind speeds reached the [26, 28] range. However, we also found that the average number of crashes per wind speed level was higher when there was no wind (Query 5).

These queries motivated us to conduct some further investigation. We noted that attributes that we were exploring (precipitation levels, average temperature, and wind speed) vary by location of the reporting weather station. For example, it might be extremely windy at only a few locations; hence, there may be less weather stations experiencing a very high wind speed. We do not want the varying number of weather stations reporting certain values to skew our analysis of the association between wind speed and number of crashes. In order to address this issue and account for a difference in the number of stations reporting, we decided to conduct three more queries. The last three queries (Query 6, Query 7, Query 8) report the average number of stations reporting certain values of each attribute (e.g. on average, how many weather stations report a precipitation level of 5 a day?) and also included the average number of crashes per day per station.

From Query 6, we can see that the average number of crashes per day per station did not vary with the changing temperature. Hence, there seems to be no association between daily average temperature and average number of crashes per day per station. From Query 7, we can see that the average number of crashes per station drastically decreased as precipitation levels increased. This implies an inversely proportional association between precipitation levels and average number of crashes. From Query 8, we can see that the average number of crashes were slightly higher (values centered around 2 crashes per day per station) for lower wind speeds of 1-15 than for higher wind speeds of 16-28 (values centered around 1.3 crashes per day per station). This implies an inversely proportional association between wind speed and average number of crashes per day per station. Thus, the results from Query 7 and Query 8 suggest that average crashes per day per station decreases as weather conditions become more extreme (higher precipitation levels or higher wind speed). We theorize that this is because more people opt to stay home as weather conditions worsen and hence, there are less vehicles on the road and thus, less crashes.

Lastly, we were curious which vehicles and factors were the most common contributors to NYC car crashes. To investigate this, we queried the top ten most common factors and vehicles (Query 9). We found that the most common vehicles involved in crashes were sedans and station wagons, and the most common contributing factors were driver inattention and following too closely. The single most likely crash scenario involved a single sedan for an unspecified reason.